

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Teknologi Informasi telah mengalami perkembangan pesat khususnya internet yang kini menjadi peranan penting. Kebutuhan *user* yang semakin beragam dalam Teknologi Informasi melahirkan sebuah cabang ilmu yaitu pencarian informasi.

Sistem pencarian data umumnya menampilkan hasil pencarian berdasarkan peringkatnya dalam daftar yang panjang, kemudian *user* diharuskan memilih data yang sesuai dengan yang mereka cari. Sayangnya sebagian besar *search engine* menggunakan paradigma tersebut. Selain itu *search engine* juga memiliki karakteristik presisi hasil pencarian yang rendah.

Search engine masih belum dapat memproses *keyword* yang *user* masukkan dengan baik. Kendala-kendala tersebut sering sekali membuat *user* mengalami kesulitan dalam menemukan informasi yang mereka cari sehingga memakan banyak waktu.

Untuk memudahkan pencarian, perlu dilakukan pengelompokan data hasil pencarian sebelum ditampilkan kepada *user*. Salah satu caranya dengan mengelompokkan data hasil pencarian menjadi kelompok-kelompok dokumen (*cluster*) berdasarkan topik masing-masing menggunakan teknik *document clustering*. Setiap kelompok dokumen diberi label yang berupa frase untuk

memberi gambaran umum mengenai dokumen-dokumen dalam kelompok tersebut.

Algoritma yang bisa menjadi solusi *document clustering* salah satunya adalah algoritma *Suffix Tree Clustering* (STC), yang merupakan algoritma yang menggunakan frase sehingga prosesnya lebih sederhana dibandingkan dengan algoritma yang lain. STC tidak memperlakukan dokumen sebagai suatu himpunan kata-kata tetapi lebih sebagai string, yaitu memanfaatkan kedekatan informasi antar kata.

Algoritma yang diperkenalkan oleh Oren Zamir ini secara teori tergolong cepat karena memiliki kompleksitas waktu (Xn). Dimana n adalah jumlah total kata dalam *snippets* atau dokumen yang digabungkan (Branson dan Greenberg, 2002:1). Selain itu, kelebihan yang dimiliki oleh *suffix tree clustering* ini adalah karena adanya *overlapping cluster* yang memperbolehkan dokumen yang sama pada kelompok *cluster* tersebut.

Berdasarkan uraian di atas, akan dilakukan penelitian dan penyusunan tugas akhir yang berjudul “**Implementasi Metode *Suffix Tree Clustering* Untuk Data Hasil *Search Engine*”**”.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, terdapat beberapa masalah yang dapat dirumuskan diantaranya adalah:

- 1) Bagaimana cara kerja metode *suffix tree clustering*?
- 2) Bagaimana mengimplementasikan metode *suffix tree clustering* untuk data hasil *search engine*?

- 3) Bagaimana sebuah pencarian informasi yang dilakukan search engine dapat memperoleh informasi yang lebih akurat dan memudahkan user dalam mengeksplorasi data?

1.3 Tujuan

Berdasarkan rumusan masalah di atas, dapat disimpulkan bahwa penelitian ini bertujuan untuk:

- 1) Melakukan kajian terhadap metode *suffix tree clustering*.
- 2) Mengimplementasikan *Suffix Tree Clustering* sebagai metode algoritma yang dapat mengelompokkan data dari hasil pencarian ke dalam bentuk grup-grup atau *cluster* berdasarkan kata atau frase yang terdapat dalam data-data tersebut. Dengan metode ini diharapkan hasil pencarian akan lebih mudah untuk ditelusuri.
- 3) Melakukan teknik document clustering yang dapat mengelompokkan dokumen-dokumen hasil pencarian kedalam cluster-cluster sehingga informasi yang diperoleh lebih akurat dan memudahkan user dalam eksplorasi data.

1.4 Batasan Masalah

Permasalahan yang ditemukan selama penelitian ini dibatasi oleh hal-hal yang tercantum berikut ini:

1. Data yang digunakan adalah dokumen berbahasa inggris.
2. Aplikasi *search engine* ini diasumsikan telah ada sebelumnya.
3. Data yang digunakan hasil *fetching* menggunakan fitur *google API*.
4. Metode yang akan diuji hanya metode *Suffix Tree Clustering*, sedangkan metode pencarian, *crawler*, dan *stemmer* tidak akan dibahas.

5. Aplikasi *search engine* yang akan diimplementasikan dengan metode *suffix tree clustering* hanya dapat memperoleh dokumen berbahasa inggris.
6. Aplikasi *document clustering* ini harus terhubung dengan internet (*online*).
7. *Tools* perancangan yang digunakan diantaranya flowchart (diagram alir) untuk menggambarkan langkah-langkah dan urutan-urutan prosedur dari suatu program, DFD (data flow diagram) untuk menggambarkan penyimpanan data dan proses yang mentransformasikan data, ERD (Entity-Relationship Diagram) untuk pemodelan data, dan kamus data untuk mendefinisikan data yang mengalir dari di system dengan lengkap.
8. *Tools* perangkat lunak yang digunakan dalam implementasi ini diantaranya visual studio 2008 dengan bahasa pemrograman C#, .NET Framework, ASP.NET, *Internet Information Service* sebagai *web server*.
9. Implementasi basis data menggunakan Microsoft Access 2007.

1.5 Metodologi Penelitian

Penelitian ini menggunakan metodologi penelitian sebagai berikut:

1. Metode pengumpulan data

Metode pengumpulan data dalam tugas akhir ini adalah sebagai berikut:

a. Studi Literatur

Pengumpulan data dilakukan dengan cara mempelajari literatur tentang penulisan dan mengenai hal-hal yang mendukung program aplikasi serta mempelajari dari sumber data yang lain seperti dari internet dan lain-lain.

a. Wawancara

Teknik pengumpulan data dengan mengadakan tanya jawab secara langsung dan bimbingan dari pembimbing.

2. Metode pengembangan perangkat lunak

Dalam proses pengembangan perangkat lunak, metode yang digunakan adalah metode pengembangan *prototype*. *Prototype* tersebut dievaluasi oleh pelanggan atau pemakai dan dipakai untuk menyaring kebutuhan pengembangan perangkat lunak. Iterasi terjadi pada saat disetel untuk memenuhi kebutuhan pelanggan dan pada saat yang sama memungkinkan pengembang untuk secara lebih baik memahami apa yang harus dilakukan. (Pressman,2002).

Tahapan-tahapan dalam *Prototyping* sebagai berikut:

a. Pengumpulan kebutuhan

Mendefinisikan format seluruh perangkat lunak, mengidentifikasi semua kebutuhan, dan garis besar sistem yang akan dibuat.

b. Membangun *prototyping*

Membangun *prototyping* dengan membuat perancangan sementara yang berfokus pada penyajian kepada *user* (misalnya dengan membuat *input* dan format *output*).

c. Evaluasi *prototyping*

Evaluasi ini dilakukan oleh *user* apakah *prototyping* yang sudah dibangun sudah sesuai dengan keinginan *user*. Jika sudah sesuai

maka langkah 4 akan diambil. Jika tidak *prototyping* direvisi dengan mengulangi langkah 1, 2, dan 3.

d. Mengkodekan sistem

Dalam tahap ini *prototyping* yang sudah di sepakati diterjemahkan ke dalam bahasa pemrograman yang sesuai

e. Menguji sistem

Setelah sistem sudah menjadi sebuah perangkat lunak yang siap pakai, maka dilakukan langkah 3 dahulu sebelum digunakan. Pengujian ini dilakukan dengan *Black Box Testing*.

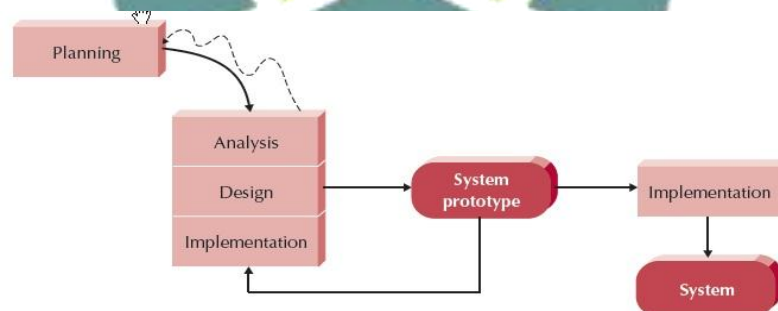
f. Evaluasi Sistem

User mengevaluasi apakah sistem yang sudah jadi sudah sesuai dengan yang diharapkan. Jika ya, langkah 7 dilakukan; jika tidak, ulangi langkah 4 dan 5.

g. Menggunakan sistem

Perangkat lunak yang telah diuji dan diterima *user* siap untuk digunakan.

Tahapan prototype ditunjukkan pada gambar 1.1 di bawah ini:



Gambar 1.1 *Prototype* (Presman, 2002)

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, metodologi penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Menjelaskan tentang teori-teori yang digunakan dalam tugas akhir ini serta untuk menjelaskan dan menyelesaikan permasalahan yang akan dikaji.

BAB III ANALISIS DAN PERANCANGAN

Bab ini juga membahas tentang rancangan aplikasi yang akan dibangun.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Membahas proses implementasi dan pengujian perangkat lunak secara detail. Proses Implementasi meliputi Persiapan *software* dan *hardware*, instalasi aplikasi, dan tampilan akhir aplikasi. Sedangkan pengujian meliputi identifikasi *software*, rencana pengujian, kasus uji dan hasil uji, evaluasi pengujian.

BAB V PENUTUP

Membahas tentang kesimpulan dan saran yang diperoleh dari penulisan tugas akhir ini.